DOCUMENT RESUME

ED 441 008 TM 030 806

AUTHOR Meloy, Linda L.; Deville, Craig; Frisbie, David
TITLE The Effect of a Reading Accommodation on Standardized Test Scores of Learning Disabled and Non Learning Disabled Students.
PUB DATE 2000-04-26
NOTE 16p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (New Orleans, LA, April 24-28, 2000).
PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Learning Disabilities; Mathematics; *Middle School Students; Middle Schools; *Reading Aloud to Others; Reading Comprehension; Sciences; *Standardized Tests
IDENTIFIERS Iowa Tests of Basic Skills; *Testing Accommodations (Disabilities)

ABSTRACT

The effect of the Read Aloud accommodation on the performances of learning disabled in reading (LD-R) and non-learning disabled (non LD) middle school students was studied using selected texts from the Iowa Tests of Basic Skills (ITBS) achievement battery. Science, Usage and Expression, Math Problem Solving and Data Interpretation, and Reading Comprehension were the four tests used. A total of 62 LD-R and 198 regular education middle school students from 2 schools participated. The most significant finding was that both LD-R and non LD students benefited from the Read Aloud test administration condition. It was expected that LD-R students would score higher in that condition, but it was not expected that non LD students would score higher, at least to the extent that they did. The Read Aloud accommodation provides a benefit to the typical student who receives it beyond the level required to alleviate the effects of a disability. The Read Aloud conditions appear to change the construct being measured for most students relative to that measured under standard conditions. The impact was greatest on the Reading Comprehension test, a test that would not be considered appropriate for such an accommodation. (Contains 3 tables and 21 references.) (SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 441 008

The Effect of a Reading Accommodation on Standardized Test Scores of Learning Disabled and Non Learning Disabled Students

A Paper Presentation for the

National Council on Measurement in Education
Annual Meeting
April 26, 2000
New Orleans, LA

By

Linda L. Meloy
Western Illinois University
and
Craig Deville and David Frisbie
The University of Iowa

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

L. Meloy

TM030806

BEST COPY AVAILABLE

Introduction

The Individuals with Disabilities Education Act Amendments (IDEA, 1997) require that students with disabilities be included in state-wide and district-wide assessments, with accommodations provided when specified by a student's IEP team. Accommodations most commonly used, as noted in the literature, have involved changes in timing or scheduling, special arrangements for the test taking setting, permitting non standard responding modes (e.g., Braille, signing), and item presentation variations such as reading aloud, reading Braille, or signing (Thurlow, Scott, and Ysseldyke, 1995). When students with disabilities take tests with such accommodations, they do so under nonstandard conditions, and consequently, the meaningfulness and appropriateness of the scores come into question (Geisinger, 1994; Phillips, 1994; Willingham, 1989). Little research has been published to date that helps us understand the effect of particular accommodations or the meanings that should be attributed to the scores when accommodations have been used.

One accommodation that is used frequently involves reading a test to a student. This is a common change used with learning disabled (LD) students who have a reading difficulty or those whose language development has been slow enough to impact their growth in preliteracy or early reading skills. The "Read Aloud" accommodation is intended to provide assistance with reading, a skill not intended to be measured by the assessment in question, so that the student can demonstrate his/her achievement without interference by a deficiency in reading comprehension. The Read Aloud should help students who have a reading deficit without giving them an advantage over those who do not receive the accommodation. Of course, the Read Aloud accommodation should not be used with a reading comprehension test or test of reading vocabulary. In such cases, reading the test to the student would drastically alter the nature of the construct being measured, an outcome that is inconsistent with the purpose of using accommodations.

Tindal and Fuchs (1999) have provided a thorough review of empirical studies that have examined the effect on test scores of altering test administration conditions. Less than one dozen studies have investigated the effects of specific accommodations, and one-third of those have looked at extended time on tests from one particular norm-referenced achievement battery, the *Iowa Tests of Basic Skills (ITBS)* (Huesman, 1999; Munger & Loyd, 1991; Perlman, Borger, Collins, Elenbogen, & Wood, 1996). Other studies have investigated response format and large print accommodations (Beattie, Grise, & Algozzine,

1983; Grise, Beattie, & Algozzine, 1982; Hollenbeck & Tindal, 1999; Tindal, Heath, Hollenbeck, Almond, & Harniss, 1999).

Five studies have investigated the oral presentation of test questions, i.e., the Read Aloud accommodation. Koretz (1997) examined Read Aloud as well as other accommodations on the *Kentucky Statewide Assessment*. He found that students with specific learning disabilities, among them reading, scored higher when read to than LD students not given that accommodation. These elevated scores were still lower than those of regular education students not given the accommodation. The author questioned the usefulness of these results due to (1) the study being an ex post facto investigation and (2) the fact that more than one accommodation was used simultaneously with some students.

Four additional studies, done by Tindal and associates in Oregon, have also looked at the use of the Read Aloud accommodation–all on that state's math assessment. Fourth grade students with IEPs and low-achieving general education students obtained higher scores than controls when the test was read to them (Tindal, Heath, Hollenbeck, Almond, & Harniss, 1999). Tindal, Anderson, Helwig, Miller, & Glasgow (1999) administered a multiple-choice math test using simplified oral language versus the standard administration of the test to middle school students with and without disabilities. In this study, no treatment effects were found. Hollenbeck, Rozek-Tedesco, Tindal, & Almond (1999) used a video read-aloud version of the multiple-choice math test with middle school students with and without disabilities, as well as a computer CD audio-only, self-paced administration. They found no significant differences across treatments or in the interaction. Finally, six math items with difficult reading loads were read to groups of fourth grade students who had been classified in terms of both reading and math achievement levels. Students with average or higher math reasoning and low reading skills improved their performance while students from other leveled math and reading skill groups did not (Helwig, Rozek-Tedesco, Heath, Tindal, & Almond, 1999).

These five studies of the Read Aloud accommodation have produced mixed results, and these results have limited generalizability for a variety of reasons, as Tindal and his colleagues have noted. They were carried out in a specific context, the *Oregon State Assessment*; the curricular area they focused on – mathematics – generally has a low reading load; and only two grade levels, fourth and eighth, were used. Furthermore, the Read Aloud procedures incorporated in two of the studies may have introduced factors

that could compromise the validity of the scores obtained. For example, in one study an overhead projector was used to show the text to students while it was being read aloud to them. This would introduce another type of reading as well as a near point-far point visual requirement unlike routine test procedures. In another study, assessment items were revised to simplify language prior to the oral reading. Such a procedure calls into question the nature of the new tasks and whether such changes alter the construct or domain being measured.

In view of the limitations of the existing studies on the Read Aloud accommodation, there is a need to examine its use more thoroughly in several subject areas and with students in various grade levels. In so doing, there is a fundamental need to establish procedures for the Read Aloud accommodation that will address the unique requirements of various subject matter areas and that could be viewed as "standard" across test administrations in which the accommodation is used. As long as the Read Aloud procedures vary from study to study, there will be little reason to compare results across studies or to establish firm recommendations for practice about the effect of the accommodation and whether its use compromises score interpretations. And because the reading demands of assessments tend to vary across subject areas, studies need to incorporate a variety of curricular areas so that such differences can be examined.

The purpose of the present study was to examine the effect of the Read Aloud accommodation on the performances of LD-R (i.e., learning disabled in reading) and non LD (regular education) middle school students using selected tests from the *ITBS* achievement battery. The specific research questions addressed by this study were:

1. Is there a difference in mean scores between LD-R and non LD students when administered each selected *ITBS* test under standard conditions?
2. Is there a difference in mean scores between LD-R and non LD students when administered each selected *ITBS* test under the Read Aloud conditions?
3. Is there a difference in mean scores of LD-R students when administered each selected *ITBS* test under the Read Aloud and standard conditions? And, is there a difference in mean scores of non LD students when administered each selected *ITBS* test under the Read Aloud and standard conditions?

4. Is there an interaction between student status as LD-R or non LD and the administration conditions for each of the selected *ITBS* tests?

Four *ITBS* tests were chosen for this study: Science, Usage and Expression, Math Problem-Solving and Data Interpretation, and Reading Comprehension. Science was chosen because its items contain a sizeable reading load, even though the reading is geared for below-average readers in the grade level of the test. The Usage and Expression test also has considerable reading, but the type of reading for some items makes the context somewhat unique. Students might need to read a paragraph and decide how to change it: replace some words, drop a sentence, reorder sentences, or add a sentence in a particular location. The Math Problem Solving and Data Interpretation test contains word problems, so there is interest in knowing whether the amount and type of reading required by those items might affect students' abilities to demonstrate their math ability. Finally, the Reading Comprehension test was included so that an estimate of the effect of reading that test aloud could be obtained. Even though a Read Aloud accommodation is inappropriate to use with this test, there is ample anecdotal evidence from educators that it is being used. There is a need to estimate what the effect of the inappropriate use of this accommodation is so that, when it is used in error, the test administrator can be helped to understand the magnitude of the distortion that would occur if the scores were interpreted as reading comprehension measures.

Middle school students were used for the study because they have been in school long enough to have been identified as learning disabled in reading, based on a significant ability-achievement discrepancy (the criterion used by the state). These students have had more time to establish records of slow progress in reading skill development relative to elementary students. In addition, students in these grades experience assessments with significant reading requirements relative to those in the earlier elementary grades.

Methodology

Subjects and Sampling

Two middle schools, having grades 6-8, from a single Midwestern school district of approximately 5285 students (K-12) participated. School #1 had 612 students, was 17.1% minority (primarily Hispanic), had 21.2% in special education, and had 29% eligible for free/reduced-price lunch. School #2 had 585 students, was 20.1 % minority (primarily Hispanic), had 18.1% in special education, and had 32% eligible for free/reduced-price lunch.

The students chosen for this study were both non LD regular education students and LD-R students. This latter group needed to meet the following criteria for inclusion in the study:

(1) currently identified as a learning disabled student and receiving special education services due to that disability;

(2) cumulative file documentation of an individually administered intelligence test score (specifically, a *Wechsler* Verbal, Performance, or Full Scale) at or above 85; and

(3) at least one reading goal on the student's current IEP.

Students with additional diagnoses or service labels such as behaviorally or emotionally disturbed (BD/ED) were not included in the study.

A total of 260 students from the two schools participated in the study, including 98 sixth graders, 84 seventh graders, and 78 eighth graders. There were 129 females and 131 males in the sample. LD-R students comprised 62 of the sample and students in regular education (non LD) numbered 198. An examination of average *ITBS* Composite test scores from the previous year, a rough indicator of overall achievement, showed scores at or slightly above the 50th percentile nationally for each of the three grade groups.

Participating students needed to be recruited for the study, and their parent/guardian needed to give permission in advance. Families of students in both schools received a letter requesting participation via mailed newsletter enclosure and mailed report card enclosure in School #1 and by student carried newsletter enclosure and student carried report card enclosure in School #2. To increase the number of special education students volunteering for participation, a separate letter was sent two weeks later to the homes of students on each school's LD roster. This letter was followed by a phone call from a special education teacher in the student's school requesting parental permission.

Procedures

Tests from Form L of the *Iowa Tests of Basic Skills* were used. This is an alternate form of Form K, the one students would be given later in the school year for their annual district-wide assessment. All tests were given on level: Level 12 for grade 6, Level 13 for grade 7, and Level 14 for grade 8. Testing was done across six school days that spanned ten calendar days in November, 1999. The standard test administrations were done in classrooms by two of the authors and by three volunteer certified staff in the

schools who were familiar with *ITBS* testing procedures. Testing groups ranged in size from 14 to 30, with an average of 21 per group. All Read Aloud testings were done in classrooms by the first author using a script that had been formulated for each test at each grade test level. These groups ranged in size from 14 to 27, with an average of 22 per group.

The scripts for the Read Aloud condition were developed in several stages. To begin, the first author and three Department of Special Education colleagues individually read the four Level 12 (Grade 6) tests, noting the time they needed to read the tests as well as the questions that arose for each reader that were idiosyncratic to each test. The main questions that arose pertaining to all tests were (method chosen in brackets after each):

1. How many times should you read an item stem? [once, unless a student asked for a repeat]
2. How many times should you read the options for each item? [once, unless a student asked for a repeat on all tests except 1 – 21 on Usage & Expression, where first the lines were read with pauses at the end and then the entire sentence read (see #1 in next paragraph)]
3. How much wait time should there be between the item stem and the first option? [regular reading flow except on Math Problem Solving and Data Interpretation, where ten-, twenty-, and thirty-second wait time was given, depending on item difficulty]
4. How much wait time should there be between the options of an item? [regular reading flow]
5. How does the reader avoid unintended voice inflections that might cue the answer? [conscious attempt NOT to do so]

Some questions that were unique to specific tests included (method chosen in brackets after each):

1. How can you most clearly read item options that are organized as continuous prose on consecutive lines? (Usage & Expression) [on 1 – 21 did not read "A., B., C…" before options, just paused; then re-read entire sentence]
2. How much of a "story teller" should the reader be when reading the reading selections? (Reading Comprehension) [passages read in "story telling style" with expression]
3. How should you read information in charts and graphs? (Math and Science) [read titles followed by bottom to top and left to right reading of chart and graph labels]

4. At what point in presenting a math problem should you pause for "thinking time", and how much thinking time should be given? (Math) [read question stem; then gave, depending on item difficulty, ten, twenty, and thirty seconds to work prior to reading the options; students allowed to ask for repeats and more time to work]

As a result of the highly consistent feedback obtained from the four faculty, the first author designed scripts for each of the four tests at each grade level. Those scripts included, in addition to the methods contained in brackets above: (1) adaptations of the directions for each test; (2) directions for how each graph and chart was to be read; and (3) numerous adaptations for the Usage & Expression test due to such things as the lettering of response options, the use of numbered sentence paragraphs, and the word-choice error section.

All but three Read Aloud administrations took more time than standard administrations (Grade 6 Math at School #2 and Grade 8 Math at both schools required the same time as the standard administration). The average Read Aloud times for the various tests at the three grade levels are shown in Table 1.

Table 1. Average Times for Read Aloud Administrations

Tests	Grade 6	Grade 7	Grade 8	Standard Time
Science	39	33	34	**30**
Usage & Expression	27	25	25	**24**
Math Problems	32	37	30	**30**
Reading Comp	48	50	50	**40**

The research design for this study incorporated both LD-R and non LD students so that the effect of the Read Aloud could be examined for both groups, and so that a possible interaction could be examined. Interest was in checking whether the accommodation worked for the LD-R students and did not work for the non LD students (Tindal, Helwig, & Hollenbeck, 1999; Fuchs, 1999). In order to equalize initial group differences as much as possible, LD-R and non LD students at each grade level were randomly assigned to one of the two test conditions, standard administration or Read Aloud, for all of their testing.

There were 127 students (49%) in the standard administration and 133 (51%) in the Read Aloud group. To permit comparisons across subject areas, each student was administered all four tests and remained in the same condition for each.

Results

Table 2 contains descriptive statistics about the performance of the LD-R and non LD students on the four *ITBS* tests for the two administration conditions. All test scores are reported on the normal curve equivalent (NCE) scale (mean of 50 and standard deviation of 21.06). NCEs were chosen for this analysis because they permit the scores of students who took different test levels within the same conditions to be combined, and, unlike percentile ranks, NCEs can be used for computational purposes without distortion.

Table 2: Descriptive Statistics by Test Administration Condition for LD and Non LD Students

Test	Admin. Condition	Student Status	N	Mean NCE	S.D.
Science	Standard	Non LD	98	57.58	23.12
		LD-R	29	30.59	10.97
	Read Aloud	Non LD	99	64.90	17.30
		LD-R	32	48.19	17.59
Usage&Exp.	Standard	Non LD	98	52.69	22.85
		LD-R	29	25.24	14.70
	Read Aloud	Non LD	100	63.79	18.02
		LD-R	33	41.97	13.02
Math	Standard	Non LD	98	57.89	20.65
		LD-R	29	32.86	16.82
	Read Aloud	Non LD	100	65.14	18.12
		LD-R	33	43.36	18.80
Reading	Standard	Non LD	98	55.35	20.53
		LD-R	29	30.28	17.51
	Read Aloud	Non LD	100	68.40	16.49
		LD-R	33	50.09	17.95

Non LD and LD-R Differences

The mean scores for the non LD students within both administration conditions were higher than those for the corresponding group of LD-R students. This outcome is consistent with the definition of learning disabilities, i.e., achievement at a lower level than expected for someone who is at least average in cognitive ability. Within the standard administration condition, the mean scores of the non LD students on all four tests were average to high average (in the range 53-68), but the mean scores for the LD students were low average to average (in the range 25-50).

Test Administration Condition Differences

The means in Table 2 indicate that both the non LD and LD-R students in this sample scored higher on all tests under the Read Aloud conditions than under standard conditions. In addition, the mean difference between conditions was larger for the LD-R students. The mean score differences for the non LD students were from 7.32 to 13.05 NCE points across the four tests, which is about one-half a standard deviation on the NCE scale. The mean differences for the LD students were from 10.50 to 19.81, or about three-fourths of a standard deviation on the NCE scale. Score variability also was different for the two conditions. For the non LD group, the standard deviation was about 25 percent smaller under the Read Aloud conditions; for the LD-R group, however, the differences were quite mixed. In Usage & Expression it was about 12 percent lower, in Reading Comprehension it was about the same, in Math Problems it was about 12 percent higher, and for Science it was 80 percent higher.

A two-way (2 x2) analysis of variance, with test administration condition and student status (non LD/LD-R) as the two fixed factors, was performed for each of the four *ITBS* tests. The results, which support the observations noted above, are shown in Table 3. For each test, Table 3 shows that the main effects for test administration conditions and student status are statistically significant ($p<.000$ level). The interaction effects, however, are not significant.

Table 3: Tests of Between Subjects Effects for Each Content Area Test

Test	Source	Sum of Sq.	df	Mean Square	F	Sig.	Eta Sq.
Science	Admin	7217.046	1	7217.046	19.468	.000	.071
	LD-R	22202.954	1	22202.954	59.892	.000	.191
	Admin*LD-R	1229.219	1	1229.219	3.316	.070	.013
	Error	94162.746	254	370.719			
	Total	937507.00	258				
Usage&Exp.	Admin	9109.35	1	9109.35	24.737	.000	.088
	LD-R	28565.862	1	28565.862	77.572	.000	.233
	Admin*LD-R	373.240	1	373.240	1.014	.315	.004
	Error	94271.686	256	368.249			
	Total	849904.00	260				
Math	Admin	3708.652	1	3708.652	10.197	.000	.038
	LD-R	25772.858	1	25772.858	70.863	.000	.217
	Admin*LD-R	124.228	1	124.228	.342	.559	.001
	Error	93106.890	256	363.699			
	Total	939197.00	260				
Reading	Admin	12711.074	1	12711.074	37.536	.000	.128
	LD-R	22141.920	1	22141.920	65.386	.000	.203
	Admin*LD-R	537.999	1	537.999	1.589	.209	.006
	Error	86690.724	256	338.636			
	Total	964131.00	260				

Discussion

The most significant finding from this study is that both LD-R and non LD students benefited from the Read Aloud test administration condition. It was expected that LD-R students would score higher with the Read Aloud, but it was not expected that non LD students would do so, at least to the extent that they did. It appears that the Read Aloud accommodation provides a benefit to the typical student who receives it beyond the level required to alleviate the effects of a disability. It is true that the average difference between conditions is greater for LD-R students, but their standard condition scores are also at a much lower level. In addition, there were no significant interaction effects found in these four testing situations.

These results are consistent with two recent accommodation research efforts. Koretz (1997) found that LD students given a Read Aloud accommodation achieved higher scores than LD students who were given a standard administration of the *Kentucky State Assessment*. Tindal, Heath, Hollenbeck, Almond, and Harniss (1999) found that all Read Aloud low-achieving and special education fourth graders in their study achieved higher on the math portion of the *Oregon State Assessment* than those in a standard

administration. However, unlike the Koretz study, the present study used only students with a learning disability in reading; and, unlike the Tindal et al. study, the present one demonstrated even stronger effects in subject areas other than mathematics. Finally, these results are consistent with those of Edwards (1970), who used a large sample of Iowa regular education students and found that those who were read the *ITBS* achieved higher scores.

But, these results do not support the accommodation position of Tindal, Helwig, and Hollenbeck (1999), who maintained that an accommodation is justified if it positively impacts the performance of students with disabilities and is neutral for non-disabled students. The Real Aloud accommodation was not neutral for this sample of non LD students. Also contained in the special education accommodation literature is Fuchs' (1999) position that an accommodation is justified if students with disabilities perform at least one standard deviation higher with such an accommodation than students without disabilities who receive the same accommodation. This LD-R sample performed only about three-fourths of a standard deviation higher under the Read Aloud conditions compared to non LD students who performed about one-half of a standard deviation higher. Therefore, the results of this study would not support the general use of the Read Aloud accommodation for students with disabilities taking standardized achievement tests.

Clearly, when average performance of groups is used to establish generalizations, exceptional cases to such generalization can be identified. For this reason, it does not seem appropriate to recommend that no student be given the Read Aloud accommodation. Certainly most LD-R students would experience improved performance with it, and some non LD students would show no improvement or fairly negligible improvement with it. What seems most clear is that the Read Aloud conditions appear to change the construct being measured for most students relative to that measured under standard conditions.

Why might the Read Aloud accommodation yield higher scores for all students – LD-R and non LD? One rather obvious explanation is that the three content area tests – Science, Usage & Expression, and Math – probably measure reading ability in addition to the content skills/knowledge they were designed to measure. Despite attempts by test developers to minimize the need for reading skills on a science test, for example, the achievement of complex ideas about designing experiments cannot be assessed well without placing some reading demands on the student. Another explanation is that reading passages to students that they would normally read themselves changes the task into a measure of listening

comprehension. It is known that listening comprehension level predicts comprehension of content better than reading skill level for many students, whether LD in reading or not (Harris & Sipay, 1985). Some other plausible explanations include the possibility that the Read Aloud procedure: (1) assisted students in maintaining their attention on individual test items and thereby reduced the chance of or need for skipping items; (2) permitted additional total test time for students to consider questions and their responses; (3) allowed those who normally are slower working students not to be bothered by the quicker pace of many of their peers; and (4) unintentionally cued students to answers based on the reader's expressive style.

The Read Aloud condition had the greatest impact on the Reading Comprehension test, a test that would not be considered appropriate for such an accommodation. That subtest was included in this study so that the effects of such an administration, done by many school personnel despite warnings of its threat to validity, could be examined. The results obtained here fully support the rational position that Read Aloud should not be used when reading is the underlying construct of interest in an assessment. Reading Comprehension tests, when read aloud, do not measure mainly reading skills, but rather a combination of student's listening skills and receptive vocabulary. Consequently, if Read Aloud is used inappropriately with a reading test, the scores should not be interpreted as indicators of reading comprehension. The meaning of such scores necessarily remains ambiguous.

Further research on the Read Aloud accommodation is certainly needed. Beyond the mere replication of this work, some additional aspects of the Read Aloud procedures should be studied, and further refinements in design and sampling would be helpful. Studies that incorporate different pacing arrangements and different expressive styles of reading would address important issues. Regarding pacing, varying approaches could be studied, such as single readings of questions/passages versus two readings and/or varying times given between question stem and response options. Oral reading varies among individuals, so different expressive reading styles, such as story-like versus computer-generated, could be investigated.

Regarding design and sampling refinements, larger samples of LD-R students are needed to establish stable estimates of the effects of the conditions, though such students are difficult to locate in one place, as well as difficult to recruit for studies. Using students from different districts could be done as long as similar criteria for classification as LD-R were used. Designs using students as their own controls

and gain scores as the main unit of analysis may be useful. If two equivalent forms of a test were available, students could be given both administration conditions, counterbalanced to reduce possible order effects.

References

Beattie, S., Grise, P., & Algozzine, B. (1963). *Effects of test modifications on the minimum competency performance of learning disabled students*. Learning Disability Quarterly, 6, 75-77.

Edwards, R.P. (1970). *A comparison between standardized and listening-reading administrations of the Iowa Tests of Basic Skills*. Unpublished doctoral dissertation, Iowa City, IA: The University of Iowa.

Fuchs, L.S. (1999). Curriculum-based measurement: Updates on its application in standards-based assessment systems. Charlotte, NC: Council of Exceptional Children.

Geisinger, K.F. (1994). *Psychometric issues in testing students with disabilities*. Applied Measurement in Education, 7(2), 121-140.

Grise, P., Beattie, S., & Algozzine, B. (1982). *Assessment of minimum competency in fifth grade learning disabled students: Test modifications make a difference*. Journal of Educational Research., 76 (1), 35-40.

Harris, A.J. & Sipay, E.R. (1985). How to increase reading ability: A guide to developmental and remedial methods. New York: Longman, Inc.

Helwig, R., Rozek-Tedesco, M., Heath, B., Tindal, G., & Almond, P. (in press). *Reading as an access to math problem solving on multiple choice test*. Journal of Educational Research.

Hollenbeck, K., Rozek-Tedesco, M., Tindal, G., & Almond, P. (1999). *A student-controlled versus teacher-controlled accommodation for large scale math tests*. Manuscript submitted for publication.

Hollenbeck, K. & Tindal, G. (1999). *Accommodating students on large scale writing tests: Consideration of time in the writing process*. Manuscript submitted for publication.

Huesman, R.L., Jr. (1999). *The validity of ITBS reading comprehension test scores of learning disabled and non learning disabled students under extended time conditions*. Unpublished doctoral dissertation, Iowa City, IA: The University of Iowa.

Individuals with Disabilities Education Act Amendments of 1997, 20 U.S.C. § 1400 et seq (1997).

Koretz, D. (July 1997). *The assessment of students with disabilities in Kentucky* (CSE Technical Report 431). Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing.

Munger, G. & Loyd, B. (1991). *Effect of speededness on test performance of handicapped and nonhandicapped examinees*. Journal of Educational Research, 85(1), 53-57.

Perlman, C., Borger, J., Collins, C., Elenbogen, J., & Wood, J. (April 1996). *The effect of extended time limits on learning disabled students' scores on standardized reading tests*. Paper presented at the annual meeting of the National Council on Measurement in Education, New York, New York.

Phillips, S.E. (1994). *High-stakes testing accommodations: Validity versus disabled rights*. Applied Measurement in Education, 7(2), 93-120..

Thurlow, M., Scott, D., & Ysseldyke, J. (1995). *A compilation of states' guidelines for accommodations in assessments for students with disabilities* (Synthesis Report 18). Minneapolis, MN: University of Minnesota National Center on Educational Outcomes.

Tindal, G., Anderson, L., Helwig, R., Miller, S., & Glasgow, A. (1999). *Accommodating students with learning disabilities on math tests using language simplification*. Unpublished manuscript.

Tindal, G. & Fuchs, L. (1999). *A summary of research on test changes: An empirical basis for defining accommodations*. Lexington, KY: Mid-South Regional Resource Center.

Tindal, G., Heath, B., Hollenbeck, K., Almond, P., & Harniss, M. (1998). *Accommodating students with disabilities on large-scale tests: An empirical study of student response and test administration demands*. Exceptional Children, 64(4), 439-450.

Tindal, G., Helwig, R., Hollenbeck, K. (1999). *An update on test accommodations*. Journal of Special Education Leadership, 12(2), 11-20.

Willingham, W.W. (1989). *Standard testing conditions and standard score meaning for handicapped examinees*. Applied Measurement in Education, 2(2), 97-103.